International Journal of Integrated Engineering:
Special Issue 2018: Data Information Engineering, Vol. 10 No. 6 (2018) p. 197-202.
© Penerbit UTHM DOI: https://doi.org/10.30880/ijie.2018.10.06.028

Review on Intrusion Detection System Based on The Goal of The Detection System

Mohammad Khamees Khaleel[1,2], Mohd Arfian Ismail[1*], Umar Yunan[3], Shahreen Kasim[4]

[1]Faculty of Computer Systems and Software Engineering, University Malaysia Pahang
[2]Department of Computer Science, College of Education, Al-Iraqia University, Baghdad, Iraq
[3]School of Industrial Engineering, Telkom University, 40257 Bandung, West Java, Indonesia
[4]Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Parit Raja, Batu Pahat, Johor, Malaysia.

Received 28 June 2018; accepted 5 August 2018, available online 24 August 2018

Abstract: An extensive review of the intrusion detection system (IDS) is presented in this paper. Previous studies review the IDS based on the approaches (algorithms) used or based on the types of the intrusion itself. The presented paper reviews the IDS based on the goal of the IDS (accuracy and time), which become the main objective of this paper. Firstly, the IDS were classified into two types based on the goal they intend to achieve. These two types of IDS were later reviewed in detail, followed by a comparison of some of the studies that have earlier been carried out on IDS. The comparison is done based on the results shown in the studies compared. The comparison shows that the studies focusing on the detection time reduce the accuracy of the detection compared to other studies.

Keywords: Intrusion Detection System, Network Security, latency in IDS, Accuracy in IDS, Computational

1. Introduction

In this era, the dependency of using the computer systems and the internet in daily life has led to serious security, privacy, and confidentiality issues due to the processes involved in the electronic transformation of data [1]. Much effort has been channeled to the improvement of the security and privacy of computer systems, however, these problems still exist in computer systems; in fact, there is no system in the world that is completely secure. Additionally, there are different types of network attacks [2]; these attacks evolve when there is a new signature with an abnormal behavior in the database of signatures. With the emergence of several types of attacks, several tools are being developed and used in several forms of network attacks. The intrusion detection system (IDS) is commonly tool used to monitor the security of network. This tool allows the monitoring of a range of network systems, cloud computing system, as well as an information system. The IDS can monitor and detect attacks which aim to compromise the security features (confidentiality, availability, and integrity) of a system. The aim of this study is to classify the IDS based on their intended goal and to compare different types of IDS in each class.

Recently, huge amounts of data are being exchanged over the network, and this has made the traditional methods of intrusion detection less efficient in performing their duties. This delay in performance simply means an ample time for the mission of the attacker to be accomplished. These reasons have made researchers focus not only on the accuracy of detection but on reducing the detection time (time needed to detect the intrusion) as well. Based on these, the studies on IDS can be grouped into two main types- those that focus on the accuracy of detection and those that focus on the time of detection and it can be seen in Fig. 1.

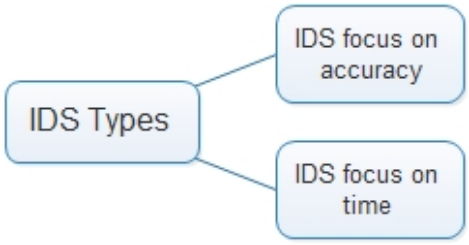

Fig. 1 IDS types

2. IDS aims to increase the accuracy of detection

There are several approaches in the studies that focus on increasing the accuracy of detection to achieve their aim, as shown in Fig. 2.

*Corresponding author: arfian@ump.edu.my

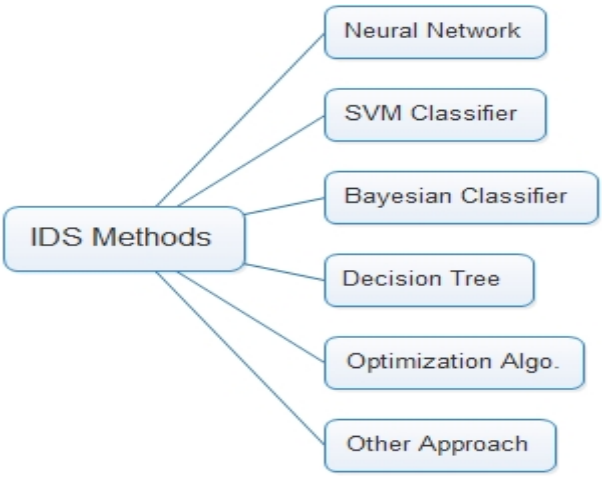

Fig. 2 IDS Methods

2.1 IDS Based on Neural Network

Neural networks (NN) were the earliest methods used for intrusion detection. Further proposals have been made for new intelligent IDS which depends on few features for intrusion detection [2]. Such systems extract data features based on the correlation and information gain concept. Firstly, the extracted data features were ranked using correlation and information gain and then were combined by using suitable method. Redundant and irrelevant data are eliminated through a pre-processing method to optimize resource commitment and minimize time complexity. In [3], the artificial neural network (ANN) in IDS. The purpose of using the ANN is to construct the classification system and then train and test the proposed system on 5 different KDD99 dataset. In another study [4], the Generalized Regression Neural Network (GRNN) model and Multilayer Perceptron Neural Network (MPNN) model were essentially discussed for host-based IDS using log files. A proposal has been made for a new fuzzy-based feature selection algorithm for the training data optimization [5] and from the study results, the boundaries between normal and abnormal class labels were evidently learned by the trained dataset. The study [5] further used random weight neural network (RWNN) as a base classifier to establish the membership vectors which corresponds to each training dataset. A backpropagation learning algorithm has been used to build an effective anomaly network IDS (ANIDS) based on the Backpropagation NN (BPNN) [6]. This new algorithm has a higher level of accuracy and detection rate while giving a low rate of false positive alarm.

2.2 IDS Based on SVM

The support vector machines (SVMs) which are also known as support vector networks are a form of supervised learning models with their corresponding learning algorithms which mainly analyze regression analysis and classification data. Several studies have deployed the SVM based IDS. An effective SVM-based ID framework for feature augmentation has been proposed by [7]. This framework provides a concise and high-quality training data for SVM classifiers using the feature-augmented technique. This does not only improve the capability of the SVM in intrusion detection, it reduces the time required for the training as well. Furthermore, a new classification framework known as GPSVM has been proposed by [8]. This framework aims to improve the rate of anomaly detection and produce a more robust level of classification accuracy using the NSL-KDD dataset without necessarily performing any form of reduction technique like feature selection or resampling. Tao, sun &suna[9] proposed a GA and SVM-based alarm intrusion detection system (FWP-SVM-GA) for application in a human-based smart IDS. The study first used a GA population search strategy, as well as the information exchange capability between individuals through the optimization of the crossover and mutation probabilities of the GA. The algorithms' convergence was improved, and the SVM training speed was enhanced. Another study [10]presented a Hypergraph-based GA (HG – GA) for intrusion detection. This framework used the same technique for feature selection and parameter setting as the SVM. An efficient IDS with a high detection rate and a low rate of false positive alarm was designed by introducing a weighted objective function. Th hyper–clique feature of hypergraph helps to minimize the complexity of the GA during its search for the best/optimal solution. Another parameter optimization algorithm based on the particle swarm algorithm (PSO–OCSVM) has been designed [11]. The study confirmed the advantages of the OCSVM in intrusion detection fields, including its strong and fast generalization capability, the simplicity of the model, the less support vector, and the great practical value.

2.3 IDS Based on Decision Tree

The decision tree (DT) is a machine learning algorithm that could be used as a classifier. Many studies in the field of IDS have been based on the DT. A novel approach for the reduction of the different representation spaces on different KDD 99 ID datasets before the application of some machine learning algorithms has been suggested [12]. The results showed the ability of the approach to reduce the training and testing times, as well as ensured a high detection rate and a low positive alarm rate when using different datasets. A novel hybrid ID method which combines an anomaly-based and misuse-based detection model (hierarchically) in a decomposition framework has been suggested [13]. First, the study used C4.5 DT to develop a misuse-based detection model before further using the model to decompose the normal training dataset into subsets. Then, an anomaly-based detection model was developed per decomposed region using the 1-class SVM (1-class SVM). An anomaly-based detection technique has also been proposed for malware detection based on the use of behavior-related information gained during malware execution on a host system [14]. Another study has also proposed Dendron, a new method for the development of DT classifiers using GA [15]. The aim of this system is to generate detection rules based on the misuse-based detection concept. This proposal provides linguistically interpretable rules to maximize security administrators benefits and to ease their tasks. A multiple-level hybrid classification model

which is a combination of DT and an enhanced fast heuristic clustering has been suggested [16]. The suggestion was found to efficiently detect intrusions with a minimal false negative alarm rate of 2.7%. It also maintained an acceptable false-alarm rate of about 9.1%. It presented a better ID performance, coupled with an appropriate balance between false negative and false positive alarm rates compared to the use of a single-level approach. A special feature of this method is its combination of both supervised (tree classifier design) and
unsupervised (clustering analysis) learning. A data mining–based framework [17] which combines two Bayesian networks and C5.0 structures has been proposed for implementation in the IDS. The system uses tree augmented Naïve Bayes and boosting method with the aim of synergistically benefiting from both methods.

2.4 IDS Based on Bayesian Classifier

In a Bayesian classifier (BC), it is believed that the function of a (natural) class is the prediction of the feature values for the class members. Several IDS studies have depended on the BC for the classification of normal and abnormal activities across a network. A clustering-based 2-stage classifier [18] has been presented for the derivation of similar subsets of system calls and models arbitrarily. A combination of clustering with supervised learning ensures the isolation of abnormal network behaviors and introduce domain-level information through carefully selected metrics. Several studies have strived to improve the detection rate of the 4 intrusion types, as well as to improve the detection accuracy of BC in the detection of R2L attacks [19]. This research achieved its objective of using BC as a data classification scheme. From the results, the BC approach performed well for R2L attack, achieving a DR of 85.35% using 3 features (23, 24 and 31), and a threshold value of 0.6. A Bayesian classifier based on the BMA of *k*-best BN classifiers called BNMA classifier for ID has also been proposed [20]. This study found the global *k*-best structures for building a BC by BMA using a DP algorithm. The study showed the BNMA classifier to have a significantly better predictive capability compared to the BN and Naive Bayes classifier which was built based on heuristic method. Even classifiers trained with smaller dataset has better performances compared to the other two classifiers trained with a larger dataset.The use of the Hidden Markov Model (HMM) for IDS has been suggested by [21]. The KDD Cup 1999 dataset for IDS was trained and tested using HMM for the applicator. Only 5 out of 41 dataset features were considered in this study. The HMM was trained for normal KDD Cup 1999 dataset's TCP connection records, while the traffic was classified as normal or attack. This showed the HMM with suitable training and parameter estimation as a powerful method for the development of IDS that can classify traffic as normal or intrusions in real-time. A two-stage clustering-based classifier has also been proposed [18] for the generation of similar subsets of system calls and arbitrarily long sequences such as Markov chains.

2.5 IDS Based on Optimization Algorithms

Optimization Algorithms such as PSO and others have been used in IDS. The S-PSO has been the most utilized technique for the construction of rule generation models [22]. The performance of the proposed model was evaluated on the KDD dataset and shown to be effective and robust. The rules generated with the model were generally better in terms of high detection accuracies, with low rates of false positive alarms. Furthermore, a MapReduce-based IDS-MRCPSO system for ID has been presented [23]. This system was developed with the aim of solving the issue of large-scale network traffic management. The system showed a parallelized efficiently with the MapReduce method. Investigations were performed on a real intrusion dataset to evaluate the system speedup, and from the outcome of the experiments, IDSMRCPSO was shown to be efficient even when increasing the size of the training dataset. The scales were proximal to the optimal speedup, hence, presented an improved detection result. A new hybrid IDS network which depends on simplified swarm optimization with weighted local search (SSO–WLS) for intrusion data classification and on intelligent dynamic swarm-based rough set (IDS-RS) for feature selection has been set forth [24]. The proposed SSO–WLS was proven to improve the anomaly detection performance based on the generation of decision rules. The method was as also tested on the KDD Cup 99 dataset for robustness. The WLS mainly aims to improve the searching process in SSO rule mining by weighing the 3 pre-determined constants. Essa [25] investigated a combined system of CFA and DT for ID feature selection. The systems' performance was also evaluated on the KDD Cup 99 data. The CFA to be deployed for feature selection was first modified before using the DT classifier to measure the generated features.

2.6 Other Approaches

Several other approaches have been used in IDS. For instance, an agent-based IDS (ABIDS) which utilizes the mechanism of agent coordination and communication has been suggested [26]. The ABIDS mainly relies on the agent-based artificial immune system (ABAIS) which is framed on the human immune systems' danger theory. The GA has also been reportedly used in another study [27] for the development of a novel GA and fuzzy logic-based anomaly detection system (NADS). The approach is in two phases which are the GA-based generation of DSNSF and the anomaly detection using a Gaussian membership function. A detailed review of IDS approaches has also been conducted [28].

3. IDS aim to Decrease the Detection Time

Detection time means the time needed to detect an intrusion. Less detection time means less time for an attacker to achieve its goal. Many methods and approaches have been used to achieve this goal. Such methods include Big Data processing tools and parallel processing concept. Principal component analysis (PCA)

was deployed to reduce the data prior to SVM parallelization and scheme implementation on the Spark platform [7]. The results showed the proposed scheme to minimize the required classification time of the classifier without a significant impact on the accuracy rate. Another study [29] has suggested a fast and efficient cybersecurity IDS framework whose performance was evaluated using machine learning algorithms and Apache Spark. This performance evaluation was carried out on the KDD'99 and NSL-KDD datasets using different feature selection and classification models. The removal of highly correlated features from the KDD'99 dataset had a minimal influence on the accuracy but minimized the required time for model training or data prediction. Finally, a comparison table as given by Table 1was developed between most types of IDS and the methods used in the literature.

Table 1: Comparison of different IDS methods

Auth.	Goal	Method	Accuracy %			
			Probe	U2R	Dos	R2L
[3]	Acc.	ANN	98.79	96.51	99.93	99.54
[4]	Acc.	GRNN	N.G	N.G	N.G	N.G
[8]	Acc.	SVM	84.27	89.28	89.79	90.72
[17]	Acc.	Bayesian	100	93.75	100	99.64
[30]	Acc.	Decision tree	99.71	66.67	99.19	89.50
[22]	Acc.	PSO	97.8	88.7	99.9	70.1
[31]	Acc.	Other approaches	73.95	82.97	100	99.55
[32]	Time	Parallel SVM on spark	94.40	96.7	90.24	89.6

Four types of intrusion are listed in table 1; the four types are Probe: Probing is an attack in which the attacker scans a machine or a networking device for vulnerabilities to be exploited to compromise the system; User to root attack (U2R): These attacks are exploitations in which the hacker impersonates a normal user account to gain access to the system and later attempts to exploit the systems' vulnerabilities to gain superuser privileges; Denial of service attacks (Dos): A DoS attack is a type of attack in which the attacker makes a computing or memory resources too busy or over-loaded to respond to legitimate user requests and hence, deny users access to a machine; Remote to user attacks (R2L): This is an attack where an attacker floods a system it cannot access with packets in order to exploit the systems' vulnerabilities on the local users' systems [33]

From table 1, we can notice that reducing the time needed to detect the intrusion (detection time) at the same time reduce the accuracy of the system for the three types of intrusion (Probe, Dos, R2L) while for U2R the accuracy is increased.

4. Summary

In the previous studies, many works have been carried out to construct the IDSbased on the intended aim to be achieved in with the selected method. The hybrid methods (more than one method) usually gives better results and accuracy. From the comparison table (Table 1), it was noticed that the accuracy of the IDS that tries to reduce the detection time was less than that of the IDS trying to increase the detection accuracy. It is, therefore, suggested that there should be a balance between the need for accuracy and the need for decreasing the detection time in the future studies by referring to various other works available such as [32]-[35]. The comparison table also shows that using optimization algorithm(PSO) to classify the intrusion is sometimes gives a better result than some classifiers like (SVM). It is, therefore, suggested that there should be a balance between the need for accuracy and the need for decreasing the detection time in the future studies by referring to various other works available such as [34]-[40].

Acknowledgement

Special thanks to the support of the sponsors from RDU Grant Vot No. RDU180307 form Universiti Malaysia Pahang.

References

[1] M. A. Mohammed, Z. H. Salih, N. Ţăpuş, and R. A. K. Hasan, "Security and accountability for sharing the data stored in the cloud," in *RoEduNet Conference: Networking in Education and Research, 2016 15th*, 2016, pp. 1-5.

[2] H. Debar, M. Dacier, and A. Wespi, "A revised taxonomy for intrusion-detection systems," in *Annales des télécommunications*, 2000, pp. 361-378.

[3] I. Manzoor and N. Kumar, "A feature reduced intrusion detection system using ANN classifier," *Expert Systems with Applications*, vol. 88, pp. 249-257, 2017.

[4] S. K. Gautam and H. Om, "Computational neural network regression model for Host based Intrusion Detection System," *Perspectives in Science*, vol. 8, pp. 93-95, 2016.

[5] R. A. R. Ashfaq, Y.-l. He, and D.-g. Chen, "Toward an efficient fuzziness based instance selection methodology for intrusion detection system," *International Journal of Machine Learning and Cybernetics*, vol. 8, pp. 1767-1776, 2017.

[6] Z. Chiba, N. Abghour, K. Moussaid, and M. Rida, "A novel architecture combined with optimal parameters for back propagation neural networks applied to anomaly network intrusion detection," *Computers & Security*, 2018.

[7] H. Wang, J. Gu, and S. Wang, "An effective intrusion detection framework based on SVM with feature augmentation," *Knowledge-Based Systems*, vol. 136, pp. 130-139, 2017.

[8] M. S. M. Pozi, M. N. Sulaiman, N. Mustapha, and T. Perumal, "Improving anomalous rare attack detection rate for intrusion detection system using support vector machine and genetic programming," *Neural Processing Letters*, vol. 44, pp. 279-290, 2016.

[9] P. Tao, Z. Sun, and Z. Suna, "An improved intrusion detection algorithm based on GA and SVM," *IEEE Access*, 2018.

[10] M. G. Raman, N. Somu, K. Kirthivasan, R. Liscano, and V. S. Sriram, "An efficient intrusion detection system based on hypergraph-Genetic algorithm for parameter optimization and feature selection in support vector machine," *Knowledge-Based Systems*, vol. 134, pp. 1-12, 2017.

[11] W. Shang, P. Zeng, M. Wan, L. Li, and P. An, "Intrusion detection algorithm based on OCSVM in industrial control system," *Security and Communication Networks*, vol. 9, pp. 1040-1049, 2016.

[12] Y. Chen, Y. Li, X.-Q. Cheng, and L. Guo, "Building Efficient Intrusion Detection Model Based on Principal Component Analysis and C4. 5," in *Communication Technology, 2006. ICCT'06. International Conference on*, 2006, pp. 1-4.

[13] G. Kim, S. Lee, and S. Kim, "A novel hybrid intrusion detection method integrating anomaly detection with misuse detection," *Expert Systems with Applications*, vol. 41, pp. 1690-1700, 2014.

[14] D. Moon, S. B. Pan, and I. Kim, "Host-based intrusion detection system for secure human-centric computing," *The Journal of Supercomputing*, vol. 72, pp. 2520-2536, 2016.

[15] D. Papamartzivanos, F. G. Mármol, and G. Kambourakis, "Dendron: Genetic trees driven rule induction for network intrusion detection systems," *Future Generation Computer Systems*, vol. 79, pp. 558-574, 2018.

[16] L. Prema Rajeswari and A. Kannan, "An intrusion detection system based on multiple level hybrid classifier using enhanced C4. 5," *Communications and Networking Madras Institute of Technology. Chennai, India: IEEE*, pp. 75-79, 2008.

[17] F. Y. Nia and M. Khalili, "An efficient modeling algorithm for intrusion detection systems using C5. 0 and Bayesian Network structures," in *Knowledge-Based Engineering and Innovation (KBEI), 2015 2nd International Conference on*, 2015, pp. 1117-1123.

[18] O. Koucham, T. Rachidi, and N. Assem, "Host intrusion detection using system call argument-based clustering combined with Bayesian classification," in *SAI Intelligent Systems Conference (IntelliSys), 2015*, 2015, pp. 1010-1016.

[19] H. Altwaijry, "Bayesian based intrusion detection system," in *IAENG Transactions on Engineering Technologies*, ed: Springer, 2013, pp. 29-44.

[20] L. Xiao, Y. Chen, and C. K. Chang, "Bayesian model averaging of bayesian network classifiers for intrusion detection," in *Computer Software and Applications Conference Workshops (COMPSACW), 2014 IEEE 38th International*, 2014, pp. 128-133.

[21] N. Devarakonda, S. Pamidi, V. V. Kumari, and A. Govardhan, "Intrusion detection system using bayesian network and hidden markov model," *Procedia Technology*, vol. 4, pp. 506-514, 2012.

[22] Z. Yi and Z. Li-Jun, "A rule generation model using S-PSO for Misuse Intrusion Detection," in *Computer Application and System Modeling (ICCASM), 2010 International Conference on*, 2010, pp. V3-418-V3-423.

[23] I. Aljarah and S. A. Ludwig, "Mapreduce intrusion detection system based on a particle swarm optimization clustering algorithm," in *Evolutionary Computation (CEC), 2013 IEEE Congress on*, 2013, pp. 955-962.

[24] Y. Y. Chung and N. Wahid, "A hybrid network intrusion detection system using simplified swarm optimization (SSO)," *Applied Soft Computing*, vol. 12, pp. 3014-3022, 2012.

[25] A. S. Eesa, Z. Orman, and A. M. A. Brifcani, "A novel feature-selection approach based on the cuttlefish optimization algorithm for intrusion detection systems," *Expert Systems with Applications*, vol. 42, pp. 2670-2679, 2015.

[26] C.-M. Ou, "Host-based intrusion detection systems adapted from agent-based artificial immune systems," *Neurocomputing*, vol. 88, pp. 78-86, 2012.

[27] A. H. Hamamoto, L. F. Carvalho, L. D. H. Sampaio, T. Abrão, and M. L. Proença Jr, "Network Anomaly Detection System using Genetic Algorithm and Fuzzy Logic," *Expert Systems with Applications*, vol. 92, pp. 390-402, 2018.

[28] M. Ahmed, A. N. Mahmood, and J. Hu, "A survey of network anomaly detection techniques," *Journal of Network and Computer Applications*, vol. 60, pp. 19-31, 2016.

[29] G. P. Gupta and M. Kulariya, "A framework for fast and efficient cyber security network

intrusion detection using apache spark," *Procedia Computer Science*, vol. 93, pp. 824-831, 2016.

[30] L. P. Rajeswari and A. Kannan, "An Intrusion Detection System based on multiple level hybrid classifier using enhanced C4. 5," in *Signal Processing, Communications and Networking, 2008. ICSCN'08. International Conference on*, 2008, pp. 75-79.

[31] B. Luo and J. Xia, "A novel intrusion detection system based on feature generation with visualization strategy," *Expert Systems with Applications*, vol. 41, pp. 4139-4147, 2014.

[32] H. Wang, Y. Xiao, and Y. Long, "Research of intrusion detection algorithm based on parallel SVM on spark," in *Electronics Information and Emergency Communication (ICEIEC), 2017 7th IEEE International Conference on*, 2017, pp. 153-156.

[33] S. Paliwal and R. Gupta, "Denial-of-service, probing & remote to user (R2L) attack detection using genetic algorithm," *International Journal of Computer Applications*, vol. 60, pp. 57-62, 2012.

[34] M.A. Ismail, V. Mezhuyev, K. Moorthy, S. Kasim, A.O. Ibrahim, "Optimisation of Biochemical Systems Production using Hybrid of Newton Method, Differential Evolution Algorithm and Cooperative Coevolution Algorithm", *Indonesian Journal of Electrical Engineering and Computer Science*, vol.8, pp. 27-35, 2017.

[35] M.A. Ismail, V. Mezhuyev, S. Deris, M.S. Mohamad, S. Kasim, R.R. Saedudin, "Multi-objective Optimization of Biochemical System Production Using an Improve Newton Competitive Differential Evolution Method", *International Journal on Advanced Science, Engineering and Information Technology*, vol.7, pp.1535-1542, 2017.

[36] M.A. Ismail, S. Deris, M.S. Mohamad, M. A. Isa, A. Abdullah, M. A. Remli. S. M. Mohi-Aldeen, "A Hybrid of Optimization Method for Multi-Objective Constraint Optimization of Biochemical System Production", *Journal of Theoretical and Applied Information Technology*, vol.81, pp. 502-513, 2015.

[37] M.A. Ismail, S. Deris, M.S. Mohamad, A. Abdullah, "A newton cooperative genetic algorithm method for In Silico optimization of metabolic pathway production", *Plos One*, vol.10, pp.e0126199, 2015.

[38] M. F. Darmawan, H. Hasan, S. Sadimon, S. Yusuf, and H. Haron, "A Hybrid Artificial Intelligent System for Age Estimation Based on Length of Left Hand Bone," *Advanced Science Letters*, vol. 24, no. 2, pp. 1047–1051, 2018.

[39] M. F. Darmawan, S. M. Yusuf, M. A. Rozi, and H. Haron, "Hybrid PSO-ANN for sex estimation based on length of left hand bone," in *2015 IEEE Student Conference on Research and Development (SCOReD)*, 2015, pp. 478–483.

[40] O.L. C Narong, C.K. Sia, S.K. Yee, P. Ong, A. Zainudin, N.H.M Nor, M.F. Hassan, "Optimization of EMI shielding effectiveness plaster mortar containing POFA using Taguchi design and Flower Pollination algorithm method", *International Journal of Integrated Engineering*, vol.10, pp.93-101, 2018.